

PRE-DEPLOYMENT COMPONENT HOSTING ENVIRONMENT ANALYZER

RELATED APPLICATIONS

- [1] This application incorporates by reference the following commonly-assigned and co-pending U.S. Patent Applications, filed on November 10, 2003: IBM Docket Number RSW9-2003-0175US1, entitled AUTOMATIC PARALLEL NON-DEPENDENT COMPONENT DEPLOYMENT; and IBM Docket Number RSW9-2003-0177US1, entitled GENERATING SUMMARIES FOR SOFTWARE COMPONENT INSTALLATION.

TECHNICAL FIELD

- [2] The present invention relates generally to the field of enterprise data systems and, in particular, to the installation of software components across enterprise resources.

BACKGROUND ART

- [3] Many computer systems a decade ago hardware on which an operating system was installed to enable software applications to be run on the hardware. Fig. 1A illustrates such a simple configuration of hardware and software. More recently, however, businesses, governments, universities and others are taking advantage of large scale networks, including intranets and the internet, to allow users located virtually anywhere to easily access applications running on machines which are also located virtually anywhere. Thus, as illustrated in Fig. 1B, additional layers are required for a user at a browser client to ultimately (but transparently) access data through server-based applications. More importantly, such enterprise computing permits combining different, often incompatible, operating systems, applications and user interfaces into the same network.
- [4] Large applications, such as application servers, may include hundreds or more individual components to install, each of which may include numerous sub-components. One example is the IBM® WebSphere® Application Server ("WAS"). In addition to the directories and files which comprise WAS, as illustrated in Fig. 2 WAS

200 also operates in conjunction with an object-oriented data base, such as IBM's DB2®-UDB 202, and an HTTP server, such as IBM's HTTP Server 204. Each of these applications comprises many components and sub-components 206. Moreover, enterprise software is frequently deployed or installed in a cluster or group of machines. Thus, when the WAS Enterprise edition is deployed, components 206 of each of the three major components (WAS 200, DB2 202 and HTTP Server 204) are installed on many machines in order to achieve a satisfactory load balancing. Heretofore, such a deployment has been a labor intensive, time consuming and error prone activity by a system administrator installing many components across many machines in a domain. And, unfortunately, heretofore, such a deployment involves installing the files sequentially, thereby adding to the time required.

[5] An additional issue is raised due to the almost infinite number of combinations of software settings and configurations on multiple hosts with multiple parameters. Such complexity makes it extremely difficult for an administrator to devise reliable test plans to insure the validity of change to software within an enterprise. Thus, seemingly harmless upgrades, patches or new software may wreak havoc on an enterprise infrastructure. Existing software may unintentionally be compromised or corrupted by additional software or software updates. It will be appreciated that such unforeseen consequences may cause part or even all of a business's enterprise system to fail. For example, a new Java Software Development Kit (SDK) is deployed each time an application, which uses Java, is deployed. Although the Java SDKs are supposed to be back-compatible they are not. Furthermore, developers commonly use both Sun and IBM Java SDKs, introducing a number of incompatibilities. That is, Java applications which were functional under SUN Java version 1.3.1, for example, might not work properly under SUN Java 1.4.1 or IBM Java 1.3.1.

[6] The Java SDK incompatibilities described above present one of the more common problems in Java 2 Platform Enterprise Edition (J2EE) environments. However, although very harmful, this is a relatively simple problem to detect. More complicated problems are presented at the operating system (OS) and compiler

levels. Frequently at the OS level there may be incompatibilities between different versions of an OS kernel and certain applications. For instance, IBM Java SDK version 1.4.1 runs only with a Linux kernel 2.2.5 or less, while the current Linux kernel on Redhat Linux is 2.5. Thus a new deployment will likely update the kernel and consequently perturb the functionality of the Java Virtual Machine (JVM) and consequently all applications that use the JVM. A similar problem might occur with OS patches.

- [7] More subtle problems may exist at the compiler level. Although different compilers use different optimization techniques, many developers are unaware of these techniques and the differences. Thus, a syntactically correct code may run differently on two compilers. For example, IBM employs the Just in Time Compilation technique (JIT) which provides an advanced optimization for the Java code. Assume that certain code reads the time, then performs some computation and finally reads the time again. When an IBM compiler is used, the time difference between the two readings will be zero, because the compiler sees no dependency between the computation and the first time reading and thus will first execute the computation. In contrast, the same piece of code will run as intended using a Sun interpreter.
- [8] Un-installation of software poses a somewhat similar problem. There are large software applications which use services from other components. For instance, WAS uses the DB2-UDB and the IBM HTTP Server. If the users decide to un-install either of the latter, WAS will no longer function. Such dependencies extend from the very high level, such as the WAS/DB2-UDB/HTTP Server example, to the finer component level, such as libraries and jar files.
- [9] While some enterprise software includes the ability to “roll back” software changes, upgrades or installations, not all enterprise software includes this function. Consequently, the responsibility to identify negative repercussions and account for a multitude of configuration scenarios rests with the software developer. It will be appreciated that developers are increasingly unable to anticipate all potential problems as software scales into enterprises and enterprises themselves increase in scale.

- [10] Still further issues arise during an installation/deployment of enterprise applications. Various people involved in installing applications and supervising their installation have differing needs during the process. For example, while a supervisor may only need a high level summary of progress, an installation administrator should be able to access detailed information on a continuous basis. However, in a large enterprise deployment, there may be an overwhelming amount of installation information available. As noted above, there may be as many as 1000 or more different components being installed. Currently, all of the information may be written to a log file, as illustrated in Fig. 3, leaving the user to decipher the contents and identify failures or other problems. Alternatively, a custom program may be written to show the progress of the installation. Such a program generally includes hard coded scripts which take time to write and must be rewritten when additional components are added. Although existing install scripts may present some screens which reflect the overall progress of installation or which provide information about the feature of the application being installed, these screens do not reflect the status of the installation of the actual components. Coupled with the long period required by the installation process, the user is left with little or no information of the actual component progress and very often has to check the functions of the underlying operating system in order to determine progress or even confirm that the installer hasn't stalled but is still proceeding.
- [11] Thus, there remains a need for an automatic pre-deployment evaluation with the capability of notifying a user of potential conflicts with existing components.

SUMMARY OF THE INVENTION

- [12] The present invention provides methods, systems, data structures and computer program products for deploying software components, including deploying components in an enterprise environment. Components previously installed and components to be installed are identified. Conflicts between such components are then identified. A user may be notified and provided with options. One option is to abort the installation. Another option is to continue the installation. If installation is

continued, an entry may be made in a log indicative of the conflict and of the continuation of the installation.

- [13] In one embodiment, a semantic model is employed which may be included in an installation package. The semantic model includes references among the components previously installed and those to be installed identifying deployment conflicts. In another embodiment, a data structure is employed which identifies deployment conflicts.
- [14] If an attempt is made to remove a component, an indication may be made of any components which depend on the component to be removed. The user may be notified and provided with options. One option is to abort the removal. Another option is to continue the removal. If the removal is continued, an entry may be made in a log indicative of the loss of dependency and of the continuation of the removal. Subsequently, if a component is installed or reinstalled, a dependency link may be created or recreated with the between the component to be installed and the dependent component.

BRIEF DESCRIPTION OF THE DRAWINGS

- [15] Referring now to the drawings in which like reference numbers represent corresponding elements throughout:
- [16] Figs. 1A and 1B illustrate past and present, respectively, hierarchies of computer systems;
- [17] Fig. 2 illustrates a hierarchy of enterprise applications and components;
- [18] Fig. 3 illustrates an exemplary log file displaying full details of an installation operation;
- [19] Fig. 4 illustrates a hosting environment agent as an intermediary between an operating system and higher level applications;
- [20] Fig. 5 illustrates a component dependency graph;
- [21] Fig. 6 illustrates grouping of similar-level components as determined from the dependency graph of Fig. 5;
- [22] Fig. 7 illustrates an exemplary parallel deployment of WAS components;
- [23] Fig. 8 illustrates an exemplary parallel deployment algorithm; and

[24] Figs. 9A and 9B illustrate an exemplary display based upon a WAS/HTTP Server/DB2 installation.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[25] In the following description, reference is made to the accompanying drawings which form a part hereof and which illustrate several implementations. It is understood that other implementations may be utilized and structural and operational changes may be made without departing from the scope of the present limitations.

[26] The present invention employs a "semantic model" described more fully in commonly assigned and co-pending U.S. Patent Serial Number ____, filed ____, IBM Disclosure RSW8-2003-0414, entitled eREGISTRY RECORDER AND ROLL BACK, hereby incorporated by reference. Such a model, generated by the developer and included in the installation package, provides a "taxonomy" of all software components of interest, such as all software which IBM, for example, produces or uses. The model comprises a set of entries for each application, component and sub-component being installed (hereinafter collectively referred to as "components"). The model includes:

- references or links among components indicating their deployment dependencies;

- entries indicating what other components are necessary for the proper operation of each component being installed; and

- entries indicating incompatibilities with other components likely to have been previously installed.

[27] More specifically, the components included in the semantic model may be at a very fine level of detail, such as jar files or libraries, or may be at a coarse level, such as enterprise applications. The relationships among these components may include (but are not limited to) the following exemplary relationships:

- "contains" in which a certain component contains sub-components without which the higher level component will not function;

“uses” in which a certain component is functional only in the presence of another component which is independent and not contained within the other (dependent) component;

“contradicts” in which a certain component may disable another component on the target along with components which have a “uses” relationship with the target component;

“equivalence” in which two components may be functionally interchangeable (i.e. Oracle and DB2 are both object-relational databases); and

“follows” in which a certain component must be installed after another.

[28] The semantic model is a data structure stored in a knowledge base (as more fully described in commonly-assigned and co-pending U.S. Patent Serial Number ____, filed _____, IBM Disclosure Number RSW8-2003-0413, entitled HOSTING ENVIRONMENT ABSTRACTION AGENTS, hereby incorporated by reference). The data structure need not be any particular structure; examples of possible structures include (but are not limited to) a flat file, a database, an object model, etc. The component semantic model is generated by the developer and may be bundled with the deployment package or accessed from a remote site during installation. In the event that deployment is to occur across domains, the model may be augmented with a list of target machines on which components will be installed.

[29] As illustrated in Fig. 4, the semantic model 400 serves as an intermediate structure between the operating system and higher level services. An “eRegistry” file stores a record of what has already been deployed while an “eReadMe” file stores a record of what is to be deployed. During an installation, an installation or configuration agent reads the eReadMe file and, after the installation is complete, updates the eRegistry.

[30] The present invention includes accessing the semantic model to obtain deployment dependency information, such as in graph format, and increasing the efficiency of a deployment by installing as many components as possible in parallel. Based on the deployment dependency information, it can be determined which components must be installed before other components. Fig. 5 illustrates the make up of such a dependency graph 500 in which “directed edges” (arrows) 510

represent dependencies among components 520 by pointing from a parent (dependent) component to a child component. The child must be installed before the parent. Fig. 6 illustrates the second step in the process, that of grouping components together having like dependency levels.

[31] More specifically, parallel installation is enabled through operating systems which support multi-threading. In order to detect the components which are suitable for parallel installation, an "acyclic directed graph" (DAG) is generated for the components which constitute an installation together with the dependency or precedence relationships among them. For instance, a deployment of WAS 5.0 involves numerous major components, five of which are: WAS, DB2, HTTP Server, Samples and Administration Tools. Each of these major components includes sub-components which in turn have further sub-components, and so on. For convenience, in Figure 7 WAS 700 is depicted with only two of the required components, DB2 710 and the HTTP Server 720. The directed edges in the figure depict dependencies among components. The numbers '1' - '5' identifying the components represent the order in which components may be installed in parallel, grouped in the manner illustrated in Fig. 6.

[32] Thus, before the WebSphere Application Server 700 itself may be installed, both DB2-UDB and the HTTP Server must first be installed. However, rather than installing the components 710 and 720 one at a time, certain of the sub-components 710 and 720 may be installed in parallel (simultaneously) in a specified order. Those components which are identified with a '1' may all be installed in parallel because they depend on no other components. The components identified with a '2' may be installed next, and in parallel with each other, because those lower level components (1) on which they depend have already been installed. Similarly, the components identified with a '3' may be installed next, and in parallel with each other, because those component on which they depend (2 and 1) have already been installed. And, finally, the WAS 700 itself may be installed. Rather than the deployment requiring eleven separate levels of component installation, only five levels are needed, a significant reduction. Fig. 8 illustrates an exemplary parallel installation algorithm which may be used to implement a parallel installation.

[33] The present invention also identifies potential component conflicts by implementing a pre-deployment hosting environment analyzer. Again the semantic model for software components is employed which captures the topology of software components at different levels of detail as well as capturing complex relationships among components. The deployed components on the target are recorded in the eRegistry. The installation is as follows: as soon as an eReadme file is available to deploy (an eReadme captures the information about the components that are to be deployed), the eRegistry is examined and the knowledge base (as more fully described in commonly-assigned and co-pending U.S. Patent Serial Number ____, filed ____, entitled OPTIMAL COMPONENT INSTALLATION) is accessed to download metadata about the relationship among the components to be installed and the components existing in the target. Next, the relationship data is analyzed so appropriate action may be taken in the event that a conflict is identified. For example, the installation may continue or the user may be alerted of the possible conflict. In the event installation continues, an entry may be recorded in a log for later reference. As soon the software is deployed on the target, the target eRegistry is updated with appropriate installation information.

[34] A complementary approach is to record on the distribution media information from the knowledge base pertinent to components to be deployed, including their relationships and the components targeted by these relationships. When such an approach is taken, the deployment target is not required to be accessible by an outside network, thus being appropriate for use in secure environments.

[35] With respect to problems which may arise when a component is un-installed, the following process may be employed. When the user decides to remove a component, the configuration management software (CMS) checks the eRegistry for any relationships involving the component to be removed. If any "uses" relationships exist, the CMS will warn the user of the consequences of the un-install action. For instance, if a user decides to remove DB2-UDB while WAS is present, CMS will warn the user that this action will disable WAS. If the user decides to continue the removal, CMS will flag WAS as being "dangling". During future installations, the CMS will examine the dangling applications for possible fixes. For example, if WAS

is dangling and the user decides to install Oracle, CMS will access the knowledge base and determine from the semantic model that Oracle is a functional equivalent to DB2-UDB which, if installed, will reestablish WAS to functionality. During the installation, the CMS will create an appropriate new link between WAS and Oracle by downloading and executing the necessary setup files from the knowledge base.

[36] The present invention further includes a process for generating installation summaries which convey varying levels of information, selectable by the user, through the use of the semantic model. As previously noted, an installation may be described, such as in an eReadMe file or a dependency graph, in terms of components to be installed and their dependencies. Components may be grouped on the basis of the number of components upon which they are dependent. Those components being dependent upon the most components being grouped at the highest (least detail) level and components being dependent upon the fewest (or no) components being grouped at the lowest (most detailed) level.

[37] The semantic model stores information about various types of dependencies. With respect to obtaining installation reports, the "contains" information is particularly relevant whereby certain components are expressed as being part of larger components. For instance, each of the major components of WAS has many other subcomponents, which, in turn, contain other subcomponents. The installation agent of the present invention accesses the semantic model and, according to the user's input, displays the requested amount of information, that is the selected granularity, about the progress of the installation. Thus, an inexperienced user may choose high level displays, displaying only the top WAS components, for example, while a system administrator may chose the lowest level of display with the finest granularity of the semantic model, such as files, libraries and jar files. The user may change the displayed level if, during installation, the user is not satisfied with the current selected level.

[38] During installation, progress may be constantly displayed, through a GUI application, by labeling the nodes (components) in the semantic model at the user's selected level of granularity. A unique indicator (such as a different color) may represent each different status, including (without limitation) "pending", "installing",

“completed” and “error”. If the installation fails, the user can visually track which particular component produced the failure as well as which components have been installed. This information will help an experienced user (viewing detailed information) determine what appropriate action to take and help a less inexperienced user (viewing less detailed information) to provide the proper information to send to a customer support facility.

[39] The report information may be displayed in a graphical, tree-like or directory-like structure in which the root component, at the highest level, represents the most important component (the WAS installation, for example). Less important components (the HTTP Server, for example) are displayed at successively lower levels. Figs. 9A and 9B illustrate an exemplary display, again based upon a WAS/HTTP Server/DB2 installation. Fig. 9A illustrates an exemplary screen showing a level of detail which might be selected by a supervisor needing only general information as components deploy. Fig. 9B illustrates an exemplary screen showing a level of detail which might be selected by an administrator needing very detailed information as components deploy.

[40] An additional feature may be included whereby, after the first installation in which a user has participated, a log is recorded of the user's selected preference indicating the level of displayed granularity. When the user participates in subsequent installations, the logged level is automatically used as the default, with the user having the opportunity to override the default.

[41] The objects of the invention have been fully realized through the embodiments disclosed herein. Those skilled in the art will appreciate that the various aspects of the invention may be achieved through different embodiments without departing from the essential function of the invention. The particular embodiments are illustrative and not meant to limit the scope of the invention as set forth in the following claims.